AMENDMENTS TO THE CLAIMS

This listing of claims replaces all prior versions, and listings, of claims in the application:

1 1. (Currently Amended) A method for generating masks for data de-duplication from entity
2 eponym data fields in a given set of data records, said data records each including an entity
3 location data field, the method comprising:
4   for each data record, splitting each entity eponym data field into a corresponding
5 prefix-suffix combinations combination, and for each prefix, tallying matches with a processor
6 computing a tally of distinct entity locations, and tallying matches of distinct entity locations
7 with a single derived suffix, and for each prefix and entity location combination, tallying the
8 processor computing a tally of distinct suffixes therefor; and
9   setting, by the processor, a threshold boundary wherein a prefix is defined as one of said
10 masks when one or more of the tallies [[is]] are indicative of different eponyms signifying a
11 single one of said entities particular entity, wherein the one mask enables a particular data record
12 to be matched to the particular entity by ignoring a portion of the particular data record.

1 2. (Currently Amended) The method as set forth in claim 1, said setting [[a]] the threshold
2 boundary further comprising:
3   setting the threshold boundary wherein the prefix is defined as the one of said masks
4 when one or more tallying results is indicative of the tallies indicate said entity eponym data field
5 including fields include variable data.

1 3. (Currently Amended) The method as set forth in claim 1, said setting [[a]] the threshold
2 boundary further comprising:
3   setting [[a]] the threshold boundary wherein [[a]] the prefix is defined as the one of said
4 masks when the tally of distinct suffixes is indicative of suffixes being information other than
5 entity identity.

3

1 4.   (Currently Amended) The method as set forth in claim 1, said setting [[a]] the threshold

2 boundary further comprising:

3   setting [[a]] the threshold boundary where a ratio of [[a]] the tally for said distinct

4 suffixes to [[a]] the tally for distinct entity locations is indicative of information other than entity

5 identity.

1 5.   (Currently Amended) The method as set forth in claim 1 further comprising:

2   applying an override function to said threshold boundary when ignore the one mask

3 based on a characteristic of [[said]] a data record is indicative of a requirement for improving

4 accuracy before a said prefix is defined as one of said masks.

1 6.   (Original) The method as set forth in claim 1 further comprising:

2   prior to said splitting, creating a reduced data records sub-set by eliminating records

3 having a unique entity eponym and entity location data pair.

1 7.   (Currently Amended) The method as set forth in claim 1 further comprising:

2   generating a display showing each data record as each derived prefix and each related

3 said entity location as a function of number unique suffixes concatenated with said each derived

4 prefix as a function of number of each related said entity location a graph having points each

5 representing a pair of a prefix and entity location as a function of a number of distinct suffixes

6 and a number of distinct entity locations.

1 8.   (Currently Amended) The method as set forth in claim 1 wherein said de-duplication [[is

2 a]] involves matching [[of]] each data record [[of]] representing a specific activity to a specific

3 known the particular entity of a plurality of known entities such that de-duplication duplication

4 of entities is minimized in a database of said plurality of known entities.

1 9. (Currently Amended) The method as set forth in claim 8 wherein said masks are

2 generated as rules for ignoring variable data portions of [[a said]] the entity eponym data [[field]]

3 fields and assigning a respective data record therefor to said database based on a non-variable

4 data portions portion of [[said]] the corresponding entity eponym data field.

1 10. (Currently Amended) The method as set forth in claim 9 further comprising:

2 maintaining said database by periodic application of said rules to a different [[said]] set of

3 data records to be added to said database.

1 11. (Original) A method for partitioning a plurality of data packets in a database such that

2 duplication of data groups is minimized, the method comprising:

3 selecting a primary identifier data field and a secondary identifier data field for each data

4 packet;

5 for all data packets having a non-unique primary identifier data field, using heuristic

6 procedures for splitting each primary identifier data into at least one prefix-suffix combination;

7 for each prefix, counting a first tally of how many distinct secondary identifier data fields

8 occurs, and counting a second tally of how many distinct secondary identifier data fields occur

9 with a single suffix, and for each prefix and each secondary identifier data field matched thereto,

10 counting a third tally of how many distinct suffixes occur;

11 based on said first tally, said second tally and said third tally generating masks

12 representative of prefixes applicable to said data packets having a non-unique primary identifier

13 data field such that application of said masks assigns data packets having a non-unique primary

14 identifier data field to associated common entities defined thereby; and

15 filing each of said data packets into a single file assigned to respective said associated

16 common entities defined.

1 12. (Cancelled)

1 13. (Original) The method as set forth in claim 11 wherein said primary identifier data field

2 is an intended unique entity name data field.

1   14.   (Original) The method as set forth in claim 11 wherein said masks are generated to

2   merge common entity name prefixes.

1   15.   (Original) The method as set forth in claim 11 wherein said secondary identifier data

2   field is a postal code data field.

1   16.   (Currently Amended) The method as set forth in claim 11 further comprising:

2   retaining said masks as rules for cleaning dirty data portions of a data field of each data

3   packet by removing variable data segments therefrom.

1   17.   (Currently Amended) A method of doing business comprising:

2   receiving, by a processor, a periodic log of transactions, each transaction being

3   represented by a data string including at least a name field and another identifier field;

4   selecting, by the processor, unique representative samples of said transactions;

5   for each of said samples, the processor dissecting each name field into derived a

6   corresponding prefix and suffix combinations combination, and for each derived prefix and each

7   prefix another another identifier combination, the processor counting [[the]] a number of distinct

8   suffixes and storing a tally therefor; and

9   generating, by the processor, a mask from a specific prefix when the specific prefix meets

10   a predefined decision criteria which is a function of said tally, wherein the mask is applicable to

11   the log of transactions to enable at least some of the data strings to be matched to a particular

12   entity name by ignoring variable portions of the at least some data strings.

1   18.   (Currently Amended) The method as set forth in claim 17 wherein for each said derived

2   prefix, counting prefix-another identifier combinations and storing a first tally therefor and

3   counting prefix-distinct another identifier combinations and storing a second tally therefor, such

4   that said predefined decision criteria is a function of said tallies.

1   19.   (Currently Amended) A computer memory ~~comprising~~ containing instructions that when

2   executed cause a computer to:

3   [[for]] store a given set of data records for a given set of entities, each of said data records

4   having discrete data fields including an entity identification field and an entity location field[[,]];

5   ~~computer code means for extracting a data pair from each of said records wherein said~~

6   ~~pair is defined as~~;

7   ~~for each data pair, computer code means for splitting~~ split each entity identification ~~data~~

8   ~~string~~ field into a ~~plurality~~ corresponding prefix-suffix ~~combinations~~ combination;

9   for each prefix, ~~computer code means for tallying matches with a~~ compute a tally of

10   distinct entity ~~location data string, and computer code means for tallying matches of each distinct~~

11   ~~entity location data string with a single derived suffix~~ locations;

12   for each prefix and entity location ~~data string~~ field combination, ~~computer code means for~~

13   ~~tallying~~ compute a tally of distinct suffixes therefor;

14   ~~computer code means for setting~~ set a threshold boundary wherein a prefix is defined as

15   one of said masks when one or more of the tallies is indicative of [[a]] different entity

16   identification ~~data string~~ strings in entity identification fields signifying a single one of said

17   entities; and

18   ~~computer code means for applying~~ apply said masks to said given set of data records such

19   that each record is assigned to a ~~single~~ corresponding one of said given entities.

1   20. – 21. (Cancelled)

1   22.   (New) The method as set forth in claim 1, wherein the data records comprise business

2   transaction records, and wherein the particular entity comprises a merchant.

1   23.   (New) The method as set forth in claim 1, further comprising applying the one mask

2   made up of the prefix to a new set of data records to assign at least some of the new set of data

3   records to the particular entity.

1   24.   (New) The method as set forth in claim 17, wherein the transactions comprise business

2   transactions, and the entity name is a name of a merchant.

1   25.   (New) The method as set forth in claim 24, further comprising the processor applying the

2   mask to the data strings to consolidate transactions associated with the merchant.